Do the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotientsahort form (AQ-S) primarily reflect general ASD traits or specific ASD traits?

Published as: Murray, A. L., McKenzie, K., Kuenssberg, R., & Booth, T. (2015). Do the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotient Short Form (AQ-S) Primarily Reflect General ASD Traits or Specific ASD Traits? A Bi-Factor Analysis. *Assessment*, 1073191115611230.

Do the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotient Short Form (AQ-S) primarily reflect general ASD traits or specific ASD traits?: A bi-factor analysis

Abstract

In the current study we fit a confirmatory bi-factor model to the items of the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotient Short Form (AQ-S) in order to assess the extent to which the items of each reflect a general versus specific factors. The models were fit in a combined sample of individuals with and without a clinical diagnosis of ASD. Results indicated that, with the exception of the Attention to Details factor in the AQ and the Numbers/patterns factors in the AQ-S, items primarily reflected a general factor. This suggests that when attempting to estimate an association between a specific symptom measured by the AQ or AQ-S and some criterion, associations will be confounded by the general factor. To resolve this, we recommend using a bi-factor measurement model or factor scores from a bi-factor measurement whenever hypotheses about specific symptoms are being assessed.

Keywords: bi-factor, autism spectrum quotient, psychometric, confounding

It is now widely accepted that autism spectrum disorders (ASD) are most appropriately described not in terms of a single unitary impairment, but in terms of multiple related traits. This multi-dimensionality of symptoms is reflected in current and historical diagnostic criteria, in psychometric measures of ASD symptoms, and in theoretical treatments of ASD and its etiology. Until recently, diagnosis of ASD was based on impairments in three domains sometimes referred to as the 'classical triad' of autism. These domains were: social, communication and restricted repetitive activities (APA, 1994). In current diagnostic criteria, the social and communication domains have been combined into a single dimension but restricted repetitive behaviours are considered sufficiently distinct to be maintained as a separate dimension (APA, 2013). Other associated features potentially include sensory symptoms; executive functioning and attention deficits; a local processing style; and a range of medical and psychological co-morbidities (Bauman, 2010; Ben-Sasson, Hen, Fluss, Cermak, Engel-Yeger & Gal 2009; Buck, Viskochil, Farley, Coon, McMahon, Morgan & Bilder, 2014; Happé, Booth, Charlton & Hughes, 2006; Happé & Frith, 2006; Matson & Shoemaker, 2009).

With regards to the core features of ASD, although the extent of their inter-relation and its cause are debated, it is generally assumed that they are all positively associated (Murray, McKenzie, Kuenssberg & O'Donnell, 2014; Williams & Bowler, 2014). To some, this covariation may be indicative of a general latent ASD factor that underlies the set of more specific symptoms associated with the disorder. For example, Mandy, Charman, Puura and Skuse (2014, p. 45) describe ASD as 'currently conceptualised as a behavioural syndrome, whereby a cluster of observable characteristics is posited as the manifestation of the latent ASD disease entity'.

It could be argued that this view is also implicit in studies which have modelled ASD symptoms and behaviours as involving some higher-order ASD factor, in addition to multiple specific first-order factors (e.g. Hoekstra et al. 2011; Kuenssberg, Murray, Booth & McKenzie, 2014). The ability to represent the covariation amongst ASD symptoms as a general latent ASD factor does not, however, suggest that this latent factor need be anything more than a statistical entity or psychometric convenience (e.g see van der Maas, Dolan, Grasman, Wicherts, Huizenga & Raijmakers, 2006) and in the current study we use the term 'general ASD factor' to mean variance shared among most or all symptoms (i.e. a general factor in a statistical but not necessarily causal sense). There is an important theoretical distinction between a statistical general factor and a reified general factor that represents the common cause of multiple symptoms. The latter interpretation requires much stronger assumptions about the causal structure underlying ASD symptoms. Nevertheless, the presence of a statistical general factor highlights the fact that whenever individual symptoms are measured, this measurement is likely to reflect not only symptom specific variance (represented by the specific factors), but also variance that is shared with other symptoms of ASD (represented by the general factor).

This covariation among ASD traits creates a challenge with respect to investigating the causes and consequences of specific ASD symptoms because when attempting to measure specific symptoms of ASD, the systematic variance in that measure may only partially reflect the specific symptom of interest. In addition, it may also reflect substantial variance due to a general ASD factor. When the shared variance among ASD symptoms is not accounted for, it can confound associations between specific symptoms and some criterion which can then reflect not only an association due the specific symptom, but also due to a general ASD factor. For some purposes, this conflation of general and specific variance may have important implications for tests of theoretical hypotheses. Many hypotheses in ASD research refer to specific symptoms and thus require an estimate that is unconfounded by general ASD variance. One example is the hypothesis that positive symptoms of schizophrenia should correlate negatively with social impairments of ASD (Russell-Smith, Maybery & Bayliss, 2011). In this case, a negative association with social impairment could be masked if the general factor of ASD is positively associated with these symptoms.

A solution which has been employed in other areas of psychopathology is to utilise a bi-factor measurement model (Reise, Morizot & Hays, 2007). An example of the bi-factor structure is shown in Figure 1. The bi-factor model represents items within a questionnaire as being influenced by a general factor (influencing all or most items) and a specific factor (influencing a specific subset of items reflecting some specific trait of interest). The specific factors are all orthogonal to the general factor and the specific factors are often also set orthogonal to one another but this is not necessary from a statistical point of view, merely conventional and often providing some interpretational benefits.

(Insert Figure 1 about here)

The bi-factor model allows an estimate of the extent to which items reflect covariation due to specific symptoms versus a general factor and by the same token helps identify items that are good measures of a specific factor versus a general factor. Bi-factor models have been used in this way in inventories measuring a range of clinical constructs such as attention-deficit-hyperactivity disorder (ADHD; Willoughby & Blanton, 2015), oppositional defiant disorder (ODD; Burke et al., 2014), and general psychopathology (e.g. Caspi et al., 2014). For example, Simms, Grös, Watson and O'Hara (2008) used a bi-factor model to assess the extent to which general and specific factors influenced responses to the Inventory of Depression and Anxiety symptoms (Watson et al., 2007). They found that the shared item variance was attributable approximately equally to a general factor and the relevant specific factors. However, it is not uncommon for a general factor to dominate even purportedly multi-dimensional inventories. For example, Watkins, Canivez, James, James and Good (2013) fit a bi-factor model to the Weschler Intelligence Scale for Children Fourth UK edition (WISC-IV: Wechsler, 2003) in a sample of children referred for the assessment of learning difficulties. They found that the general factor accounted for almost two thirds of the common item variance.

To date few studies have addressed the question of the extent to which the items in commonly used inventories in ASD research reflect general versus specific factors. Snow, Lecavalier and Houts (2009) fit a bi-factor model to the items of the Autism Diagnostic Interview Revised (ADI-R; Lord, Rutter & Le Couteur, 1994) in individuals with a clinical diagnosis of pervasive developmental disorder (PDD) but did not interpret the solution, reporting that it did not yield better fit than an oblique model. Lecavalier, Gadow, DeVincent, Houts and Edwards (2009), for the same reason, did not report the solution for a bi-factor model fit to the items of the Early Childhood Inventory-4 (ECI-4; Gadow & Sprafkin, 2000) and the Childhood Symptom Inventory-4 (CSI-4; Gadow & Sprafkin, 2000) in a sample of children referred to a developmental disability clinic.

However, one possibility is that in using only clinically diagnosed or referred individuals, item inter-correlations were attenuated due to range restriction, reducing the variance common to all items and thus any support for a general factor. Another study, which used a general population sample including individuals with a self-reported diagnosis of ASD, reported that a bi-factor model fit best among those assessed (Posserud, Breivik, Gillberg & Lundervold, 2013). It was fit to a 7-item screening tool, the 'autism self-report for adolescents and adults' (ASSERT), derived from Asperger syndrome (and high functioning autism) diagnostic interview (ASDI; Gillberg, Gillberg, Råstam & Wentz, 2001). Although it was the best fitting model, several core symptoms had only low loadings on the general factor.

In the current study we fit a bi-factor confirmatory factor model in a popular measure of ASD traits: the autism spectrum quotient (AQ; Baron-Cohen et al. 2001) as well as its abridged version, the autism spectrum quotient short form (AQ-S; Hoekstra et al., 2011) and evaluate the extent to which the specific symptom areas measured by each reflect the intended specific factors versus a general factor of ASD.

Method

Participants

Data came from four (1 clinical sample, 3 control samples) sources which have been separately utilised in previous studies but were combined into a single dataset for the current study. From the original samples, cases were removed if a) were under the age of 18, b) they had no responses to any of the 50 AQ items or c) if they were participants recruited for a control sample and self-reported a diagnosis of ASD or intellectual disability. This resulted in a total sample of 562 respondents (204 male; 357 female; 1 not reported) with a mean age of 30.6 years (SD=11.8 years) ranging from 18 to 69. Specific details for each of the 4 constituent samples of are provided below.

Clinical Sample

One hundred and forty seven participants were drawn from a clinical sample. These individuals all had a clinical diagnosis of either Asperger's syndrome (AS) or high functioning autism (HFA). AS was defined as meeting the criteria for HFA but with no history of delay in language while HFA was defined as meeting the criteria for autism but with normal intellectual functioning. Data were obtained from case notes from a specialist Regional ASD Consultancy service and clinical psychology services in Scotland. Diagnoses were made with reference to *Diagnostic and Statistical Manual of Mental Disorders (4th ed. Text rev.; DSM-IV-TR)* criteria by an experienced clinician and utilised clinical interview, informant interview where available and individual assessment such as neuropsychological assessment where indicated. Prior to finalising a diagnosis, each case was discussed at a multidisciplinary clinic. Participants were primarily male (n=107, 73%) and had a mean age of 33.5 years (SD 10.7 years; range = 18 to 62 years). The sample has been utilised in several previous studies (Booth, Murray, McKenzie, Kuenssberg & O'Donnell, 2013; Murray, Booth et al. 2014; Kuenssberg, Murray, Booth & McKenzie, 2014; Murray, McKenzie et al. 2014).

Control Samples

One hundred and sixty four participants came from a previous psychometric study of the AQ (Murray, Booth, McKenzie, Kuenssberg & O'Donnell. 2014). These participants were primarily female (n=125, 76%). They had a mean age of 30.1 (SD=11.3) ranging from 18 to 65.

Data for 97 participants came from an ongoing study of emotion recognition and ASD traits which have been previously utilised in a study by Murray, McKenzie et al. (2014). These participants were recruited online and from the university community and included 27 males, 69 females and 1 participant who described their gender as 'other'. The mean age of the sample was 31.1 (SD = 12.5, range = 18 to 68).

Lastly, 154 participants came from an ongoing study of sex differences in ASD traits in the general population. These have also been previously utilised in the study by Murray, McKenzie et al. (2014). These participants were recruited online and were composed of 123 females and 31 males. The mean age of the sample was 28.0 (SD = 12.4, range = 18 to 69).

Ethical approval was obtained from the relevant ethics committee in all cases. We analysed data from both clinically diagnosed individuals and individuals with no clinical diagnosis of ASD together because the development of the AQ was based on the premise that it is possible to measure ASD traits on a continuum from very extreme levels to levels found in the general population, with clinically diagnosed individuals merely being at the extreme end of this continuum. Indeed, in practice the AQ and its derivatives are utilised in both clinical and general population samples. Furthermore, restricting analyses to either those who have a clinical diagnosis of ASD or those who do not could lead to a downward bias in the parameters of a bi-factor model due to variance restriction (Murray, McKenzie et al., 2014).

Measures

AQ

The autism spectrum quotient (AQ) is a 50 item questionnaire developed to measure ASD traits in adults of normal intellectual ability (Baron-Cohen et al., 2001). The areas measured by the inventory were chosen based on the classical triad of ASD, as well as commonly associated features. The AQ is organised into 5 subscales, each with 10 items. These subscales aim to measure ‘Social Skills’, ‘Attention Switching’, ‘Attention to Detail’, ‘Imagination’, and ‘Communication’. Abbreviated item contents are provided in Table 1. The phrasing of items is such that they reflect behavioural tendencies and preferences rather than symptoms or impairments. Thus, it is most accurately characterised as a measure of autistic traits.

In this study, respondents were offered a four point response scale from 'Strongly Agree' to 'Strongly Disagree'. Some items are phrased in a 'forward' direction where choosing 'Strongly Agree' suggests higher levels of autistic traits and some are phrased in a 'reverse' direction where selecting 'Strongly Disagree' suggests higher levels of autistic traits. For this study, all items were (re-) coded in the direction of higher scores reflecting higher autistic trait levels. Many previous studies have collapsed the 4 point scale of the items into a dichotomous '0' vs '1' scoring scheme, however, this generally entails a loss of information and in the current study we scored the items using a '1' to '4' scoring scheme to maximise reliability. All participants in the current sample were administered the full AQ. To disambiguate the AQ from its short form AQ-S in the current study, we henceforth refer to the full 50 item AQ as the 'AQ50'.

AQ-S

We also present analyses of the abridged version of the AQ50, the AQ-S (Hoekstra et al., 2011). The AQ-S contains a subset of 28 items judged to provide the best measures of ASD traits from the AQ50 whilst still ensuring that all key content areas are adequately represented. The process by which the items of the AQ-S were selected from the full AQ50 is described in full in Hoekstra et al. (2011). Briefly, in a sample of Dutch controls (no ASD diagnosis), exploratory factor analyses were conducted and potentially problematic items identified on the basis of content (e.g. an item very similar in content or phrasing to another item) and the factor analysis results (e.g. low loadings on the relevant factor). After selecting an optimal factor structure, a number of potentially problematic items were excluded. A series of confirmatory factor analyses guided further refinements. Finally, confirmatory factor analyses in two independent samples (a further sample of Dutch controls and a sample of English controls) were used to verify that the factor structure developed in the initial sample provided a good representation of the item covariances in these new samples, supporting its generalisability.

The set of items selected using these methods comprised 28 of the original 50 and could be represented in terms of five specific dimensions of 'Social Skills', 'Routine', 'Switching', 'Imagination', and 'Numbers/patterns'. A higher-order 'Social Behaviour' dimension underlying all of these dimensions, except 'Numbers and Patterns,' was also supported. The 'Numbers/patterns' was relatively independent of all other dimensions and correlated with the 'Social Behaviour' factor at only r=.2.

Statistical Procedure

We estimated bi-factor confirmatory factor models based on previous psychometric evidence on the structure of the AQ50 and the AQ-S. Specifically, we estimate models closely equivalent to the structures presented in Baron-Cohen et al. (2001) for the AQ50, and Hoekstra et al. (2011) for the AQ-S.

Baron-Cohen et al. (2001) originally developed the AQ50 and suggested a 5 factor structure with ten items measuring each factor. The factors were labelled Social Skills (items 1, 11, 13, 15, 22, 36, 44, 45, 47, 48), Attention Switching (items 2, 4, 10, 16, 25, 32, 34, 37, 43, 46), Attention to Detail (items 5, 6, 9, 12, 19, 23, 28, 29, 30, 49), Communication (items 7, 17, 18, 26, 27, 31, 33, 35, 38, 39) and Imagination (items 3, 8, 14, 20, 21, 24, 40, 41, 42, 50). For the AQ50, we specified a bi-factor model with a general factor loading on all items, and five specific factors with item loadings as described above.

Hoekstra et al. (2011) developed a higher-order model for the AQ-S in which a second-order Social Behaviour factor underlay four specific factors labelled Social Skills (items 1, 11, 13, 15, 22, 44, 46, 47), Routine (items 2, 25, 34, 46), Switching (items 4, 10, 32, 37), and Imagination (items 3, 8, 14, 20, 36, 42, 45, 50). In addition, the general Social Behaviour factor was allowed to correlate with a specific Numbers/patterns factor (items 6, 9, 19, 23, 41). We evaluated a bi-factor structure based on, but not corresponding exactly to, this model. Specifically, we specified a model in which all items loaded on both a general and specific factor and these factors were mutually uncorrelated. The Hoekstra et al. (2011) study would imply that the items from the Numbers/patterns factor should not load on the general factor. In this study, we allowed them to load on the general factor so that we could evaluate all items for general versus specific factor variance.

It is common practice to compare the fit of a bi-factor model to a higher-order model, however, in the context of the current study where our goal was to estimate the extent to which AQ50 and AQ-S items reflect general versus specific factor variance, a higher-order model would provide no additional information. Further, previous studies have suggested that whenever there are unmodelled complexities in a psychometric model (e.g. cross-loadings or correlated residuals), these kinds of model comparisons may be biased in favour of the bi-factor model which, being the more general model, is relatively less sensitive to mis-specification in terms of its impact on model fit (Murray & Johnson, 2013). However, we did compare the fit of the bi-factor model with a correlated first order factor model which includes *only* specific factors and no general factor. This provided a test of whether a general factor was supported.

Models were estimated using weighted least squares mean and variances (WLSMV) estimation in *Mplus 6.11* (Muthén & Muthén, 1998-2013). WLSMV was used to account for the ordered-categorical response format of the scale. By default, this method uses pairwise deletion to deal with missingness and given the low rate (0.54%) of missing item responses and high (>98% in all cases) covariance coverage, this was judged a reasonable strategy. Latent factor variances were fixed to 1 for scaling and identification purposes.

Model fit was evaluated based on the comparative fit indices (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and the Weighted Root Mean Square Residual (WRMSR). For the CFI and TLI, values of > .90 to .95 and for RMSEA values, of <.08 were taken as indicative of good fit (Hu & Bentler, 1999; Schermelleh-Engel et al., 2003).

Strength of general and specific factors

As an index of the extent to which the items reflected a general factor versus specific factors, we computed the explained common variance (ECV) statistic (see Reise, 2012). This is computed as a ratio of the variance explained by the general factor to the variance explained by the general plus specific factors. Higher values of ECV suggest that a higher proportion of the common items variance is inventory-wide, rather than specific to a set of items in a sub-scale.

We also estimated the reliability of the individual sub-scales both before and after controlling for the general factor. Similarly, we estimated both the overall reliability of the total scales and the reliability of the total scales that were attributable to the general factor alone (i.e. controlling for the specific factors). All of these reliability indices are variants on McDonald's (1999) ω statistic, which estimates the proportion of variance in a scale (or subscale) that is attributable to a given factor or set of factors. The denominator is the total scale or relevant subscale variance and the numerator is the variance due to the relevant factor(s). For example, a measure of the reliability of a scale due to the general factor is

$$\omega_h = \frac{(\sum \lambda_{iG})^2}{(\sum \lambda_{iG})^2 + (\sum \lambda_{iS1})^2 + (\sum \lambda_{iS2})^2 + (\sum \lambda_{iS3})^2 + \cdots (\sum \lambda_{iSP})^2 + \sum {\theta_i}^2},$$

where λ_{iG} represents general factor loading for item i , λ_{iS1}…λ_{iSP}, represent the specific factor loadings for item *i* on specific factors 1-P, and ${\theta_i}^2$ represents the error variance for item *i.* When referring to scale or subscale reliability due to both the general factor and the relevant specific factor we use 'omega total' or ω_T and when referring to scale or subscale reliability due only to the general factor (for total scales) or the relevant specific factor (for subscales) we use 'omega hierarchical' or ω_h. Both ECV and the ω indices were computed from the parameter estimates obtained from the bi-factor model.

Practical implications

Although CFA analyses and indices derived from these analyses provide valuable information about the extent to which specific items and tests as a whole reflect general versus specific factors, they do not necessarily lead directly to recommendations about whether individual subscale or total scores are appropriate for use in practical applications. We, therefore, estimated additional indices to inform practical recommendations (Reise et al, 2013).

First, we estimated a measure of worst-split half reliability, β (Revelle, 1979) for the total scales and all specific subscales. Revelle (1979) has argued that lower-level subscales are usefully combined into larger higher-level scales when doing so increases β. Therefore, comparing β for the individual subscales of the AQ50 and AQ-S and the inventories as a whole can help determine the level of aggregation of items that is most appropriate. The β index was computed from the original item covariance matrix for each inventory using the 'psych' package in R statistical software (Revelle, 2013; R Core Team, 2014).

Second, we evaluated whether use of subscale scores would be appropriate using the methodology proposed by Haberman (2008). Haberman (2008) noted that a scale total score is sometimes a better estimator of the true score for a construct measured by a subscale than is the subscale score itself. This can happen if the subscales contributing to a total score are highly correlated and the reliability of the total score sufficiently exceeds that of the subscale. In this situation, using subscale scores is not recommended. To assess whether this is the case for any of the AQ50 and AQ-S subscales, the proportional reduction in mean square error (PRMSE) with respect to the construct measured by the relevant subscale was computed for both the total score and the subscale score. If the former exceeds the latter, then subscales scores should not be used. PRMSE values were computed using the 'sirt' package in R statistical software (Robitzsch, 2015).

Lastly, to assess the potential impact of confounding of specific factor associations with the general factor, we evaluated the sex difference in each of the specific factors of the AQ50 and AQ-S both before and after controlling for the general factor. We did this by estimating the association of the specific factors with sex in a first-order oblique measurement model that included no general factor to give 'confounded estimates'. We then obtained 'unconfounded estimates' by estimating the same associations using a bi-factor measurement model which controls for the general factor. A comparison of the estimates for the corresponding specific factors across these two kinds of measurement models provided a gauge of the extent to which sex differences in specific ASD symptoms are affected by whether general covariation between ASD symptoms is taken into account or not.

Results

AQ50

Model fit for the bi-factor model was good according to all model fit indices (χ^2=3646.24 (1125), p<.001; RMSEA=.06; WRMR = 1.72; CFI=.92; TLI=.91) and better than that of the first order correlated factors model (χ^2=4744.05 (1165), p<.001; RMSEA=.07; WRMR = 2.11; CFI=.88; TLI=.88) which would be considered to provide poor fit according to the CFI and TLI. These results supported the inclusion of the general factor in the AQ50.

Standardized factor loadings, total item variance explained, and the proportion of total variance attributable to the general factor for the AQ50 are provided in Table 1.

(Insert Table 1 about here)

General factor loadings ranged from |.05| to |.88|, with the general factor accounting for between .2% and 100% of the total explained item variance. Two items (29 and 49) had non-significant general factor loadings and the general factor loadings for two items (30 and 49) were negative, despite all items being coded in the same direction.

Specific factor loadings were quite variable across items. For four of the five specific factors, there was at least one item which had either a non-significant, or negative loading. Only for the Attention to Detail specific factor did all items have significant loadings in the expected direction. The items in this factor had some of the lowest general factor loadings, suggesting they capture variability that is somewhat independent of that in the remainder of the AQ50.

These patterns of factor loadings are reflected in the factor reliabilities as estimated by the ω values. As can be seen from Table 2, the general factor showed very good reliability according to the ω_h(.91), but the values for four of the specific factors were very low, ranging from .05 to .24. Attention to Details, however, displayed moderate unique reliability (.67). Overall, the ECV value (.70) suggested that most of the variance shared among items was inventory-wide rather than subscale specific.

(Insert Table 2 about here)

As the items in the Attention to Details subscale appeared to be relatively independent of the remaining items of the AQ50, an estimate of worst split half reliability for the AQ50 total score was obtained both with ('Total Score') and without these items ('Modified Total Score'). Based on a comparison of β values for the AQ50 total scales and the subscales and Revelle's (1979) suggestion that a good heuristic for when scales should be aggregated is when it results in an increase in this index, the appropriate level of aggregation would appear to be the total scale level. This was true irrespective of whether the Attention to Detail items were included in the total scale or not but more so when they were not.

Again, given that the Attention to Detail items did not appear to be strongly related to a general factor in the AQ50, we also computed PRMSE for each subscale based on a total score both including and excluding these items, however, this made little difference to results. For the Social Skills, Attention to Detail, Attention Switching and Imagination subscales, subscale score PRMSE was greater than or effectively equal to total score PRMSE, suggesting that the use of subscale scores is supported for these constructs. For the Communication subscale, however, total score

PRMSE was greater than subscale PRMSE, suggesting that this construct may be better predicted using AQ50 scores than subscale scores.

Taking all results together, with the exception of the Communication subscale, using both subscale and total scores appears to be appropriate in the AQ50. A modified total score which excludes the Attention to Detail items should be preferred to a total score of all 50 items because these items appear to reflect a relatively distinct construct compared with the rest of the items of the AQ. In addition, as the ω_h values show, subscale scores should be used in the knowledge that much of the reliability of these subscales is due to a general factor, rather than subscale specific factors.

AQ-S

The initial model specification for the AQ-S yielded a Heywood case: a negative residual variance estimate for item 32. We dealt with this by constraining this residual variance to a small positive value (.01). Parameter estimates suggested that this item was strongly related to both a general and specific factor. Inspecting these parameters and the content of the item (referring to multi-tasking ability) suggested that there was no reason not to retain the item in spite of it initially yielding an out of range parameter estimate. The model with this additional constraint showed good to excellent model fit according to all indices (χ^2= 973.13 (321), p< .001; RMSEA=.06; WRMR = 1.27; CFI=.96; TLI=.95). Model fit for the first order correlated factors model was reasonable (χ^2=1537.95 (339), p<.001; RMSEA=.08; WRMR = 1.67; CFI=.93; TLI=.92). Whilst the raw fit may suggest this model to be plausible by the cut-off values highlighted above, the decline in fit between the bi-factor and the correlated factors model was reasonably large (ΔRMSEA = .02; ΔCFI

= -.03; ΔTLI = -.03). Again, these results support the inclusion of the general factor in the AQ-S.

Standardized factor loadings, total item variance explained, and the proportion of total variance attributable to the general factor for the AQ-S are displayed in Table 3.

(Insert Table 3 about here)

General factor loadings for the AQ-S items were all significant and positive and ranged from .26 to .81. The general factor accounted for between 13% and 97% of the total explained item variance. The ECV for the AQ-S was .63.

As with the AQ50, there was some degree of variability in the magnitudes of the specific factor loadings in the AQ-S; however for the AQ-S, all specific factor loadings were a) significant, and b) positive, as would be expected. This pattern resulted in ω_h values for the specific factors ranging from .18 to .62 (see Table 2). Again, the general factor showed excellent reliability ω_h =86.

Analogous to the AQ50, we computed worst split half reliability based on a total score of the 28 AQ-S items and a modified total score which excluded the low general factor-loading Numbers/patterns items. The resulting β values suggested that the appropriate level of aggregation was the total scale level excluding the Numbers/patterns items.

The PRMSE values suggested that for the Social Skills, Imagination and Numbers/patterns subscales, the subscale score provided the best prediction for the subscale constructs. In the case of Switching, the PRMSE values were identical, and for Routine, PRMSE was slightly higher for the total score.

Collectively, the results from the AQ-S mirror the conclusions of the AQ50. The ω_h estimates suggest that much of the subscale variance is in fact attributable to the general factor, and based on the β estimates, the general factor appears to be the best level of aggregation. However, the PRMSE analysis indicated that the subscale scores remain plausible.

Criterion associations

Table 4 contains the standardized parameters for the covariate effects of sex on the specific factors of the AQ50 and AQ-S taken from the correlated traits and bi-factor models.

(Insert Table 4 about here)

As would be expected, sex had a significant effect (p<.01) on all specific factors in both the correlated traits and bi-factor models. The direction of these effects indicated that males score higher on each of these factors. For the AQ50, the confidence intervals for the covariate effect on four of the factors in the models, with and without controlling for the general factor, overlapped, suggesting that this did not significantly affect these associations. However, in the case of Attention Switching, the 95% confidence intervals did not overlap, suggesting a significant difference in the magnitude of the estimates across models. When the general factor was controlled for, the sex difference on Attention Switching increased. A similar pattern was observed in the AQ-S with the subscales of Routine, Switching and Imagination.

Discussion

In the current study, we used a bi-factor confirmatory factor model to assess the extent to which the items of the full AQ (AQ50) and its short form, the AQ-S, reflect a general factor of ASD versus factors reflecting more specific symptoms. For the AQ50, the majority of shared item variance was attributable to the general factor. The main exceptions were the items of the Attention to Details subscale which tended to have lower general factor loadings but higher specific factor loadings. Similarly, for the AQ-S, with the exception of the items in the Numbers/patterns scale, item covariance primarily reflected a general factor rather than specific factors. Overall, in both inventories, shared item variance was to a large extent inventory-wide rather than subscale-specific.

Several authors have previously discussed the implications of the extent to which items in an inventory reflect general factor as opposed to specific factor variance. One implication concerns the appropriate level of analysis, in particular, whether the strength of the general factor is such that items from different subscales should be combined into a single scale. Revelle (1979) argued that combining items from scales would be justified if it resulted in an increase in the worst split half reliability, β. For the AQ50, this heuristic suggested it should be scored in two parts: a modified scale which includes all AQ50 items except those from the Attention to Details subscale and a separate Attention to Details subscale. On the other hand, the PRMSE values suggested that if subscale scores were desired, it would generally be appropriate to use subscale scores, except for the Communication construct, which is likely to be better estimated using the AQ50 total score.

For the AQ-S, Revelle's heuristic suggested that items should be organised into a general scale comprising the items of the Social Skills, Routine, Switching and Imagination items and a separate Numbers/Patterns subscale. However, the PRMSE values indicated that only the Attention Switching and Routine constructs would be expected to be better predicted by the AQ-S general scale score than by their respective subscale scores. The use of subscale scores were supported for the Social Skills, Numbers/patterns and Imagination constructs.

If using subscale scores from the AQ50 or AQ-S, it is important to note that a large part of their variance and reliability is due to a general factor. Reise, Moore and

Haviland (2010) noted that using subscales as measures of specific factors may be misleading when their systematic variance is mostly due to a general factor. Similarly, DeMars (2013) noted that the subscales may appear to be highly reliable but will not show distinct correlations with external criteria because of confounding with the general factor. Therefore, it will be difficult to establish discriminant validity of specific symptom subscales as well as specific correlates of the symptoms they are assumed to measure. These kinds of considerations apply equally when the general factor is of interest. For example, the presence of a sub-scale structure when left unmodelled creates issues such as violations of local independence and attendant over-estimates of the reliability of the scale as a measure of a general ASD factor (e.g. Braeken,2011).

In the current study, the associations between several specific factors and sex became magnified when controlling for the general factor. In the AQ50, the sex difference on Attention Switching significantly increased after controlling for the general factor and in the AQ-S the sex differences on Routine, Switching and Imagination increased significantly after controlling for the general factor. This demonstrates that the decision to control for a general factor or not could in some cases affect substantive conclusions. In the particular case of sex differences, the most likely explanation is that controlling for the general factor reveals normative sex differences, which are otherwise conflated with the extent to which individuals exhibit general autistic-like tendencies.

One potential solution to confounding with the general or specific factors is to utilise a bi-factor measurement model for ASD inventories to obtain estimates of reliability and correlations with external criteria for the specific and general factors. Less ideal but also defensible is to use factor scores estimated from the bi-factor model that represent scores on specific symptoms controlling for general ASD variance, as well as general ASD scores controlling for specific symptoms (DeMars, 2013). These approaches may be useful in the context of testing theories which predict a correlation between a *specific* ASD symptom and some external variable. Such specific theoretical mappings have, for example, been drawn between possible emotion recognition deficits and social symptoms (Dawson, Webb & McPartland, 2005) and between restrictive repetitive activities and frontal lobe functions (Lopez, Lincoln, Ozonoff & Lai, 2005). These associations will be more difficult to assess when items reflect not only the specific symptom with which the hypothesis is concerned, but also inventory-wide variance. In this situation it will not be possible to be sure that an observed correlation is due only to the specific symptom of interest. Similarly, when a negative association between a specific symptom and some external criterion is predicted but the association with the general factor of ASD is positive, that negative association may be masked.

In some contexts, it will not be necessary or appropriate to attempt to separate out general and specific ASD variance. One example is when the test is being used for prediction (Revelle & Zinbarg, 2009). For example, when screening for or diagnosing ASD (i.e. predicting ASD status) where the goal is simply to identify an individual with overall high levels of ASD symptoms, it may matter little if item scores reflect both general and specific traits. Here an amalgamated estimate will generally identify individuals who may meet diagnostic criteria for ASD irrespective of the factor structure thought to be the best representation of the inventory. If the underlying factor structure is not relevant, the AQ-S total score may be preferred over the full AQ50 because, in spite of being 22 items longer, the AQ50 had a total score reliability that was only marginally higher than that of the AQ-S. Similarly, the

Numbers/patterns items could be omitted without adverse impact on total score reliability. Thus, participant fatigue can be reduced by administering a briefer inventory with little detriment to the reliability achieved.

As noted above, there were some exceptions to the overall tendency for the general factor to dominate item variance. In the AQ50, the Attention to Detail items were an exception to this trend, as were the Numbers/patterns items in the AQ-S. In fact, these are essentially the same items. That these items reflected a large proportion of shared variance independent of the general factor is consistent with previous research that has suggested that the items measure a construct relatively distinct from the other constructs captured by the items of the AQ (e.g. Hoekstra et al., 2011; Stewart & Austin, 2009). For example, Hoekstra et al. (2011) found that it was possible to include a higher-order factor in a CFA model of the AQ-S but the Numbers/patterns did not fit within this factor and the factor correlation between the Numbers/patterns factor and the higher-order ASD factor was only .2. Thus, evidence is accumulating that the items measured by the Numbers/patterns factor reflect a relatively distinct attribute which may not represent a core feature of ASD.

Limitations and Future Directions

In terms of study limitations, while we were able to include individuals with a very broad range of ASD traits levels (from clinically diagnosed to non-clinical levels) our sample was not population representative. Individuals with low intellectual functioning were not represented in the current sample, as the AQ was designed for individuals of normal intellectual ability. In addition, we had only self-report measures which is sub-optimal given the possibility that the autistic traits may be associated with accuracy of self-reports (Johnson, Filliter & Murphy, 2009).

In addition, our methodology was unable to probe the cause of inventory-wide variance. While it could reflect a causal general ASD factor, it could also reflect local interactions between symptoms or a range of other causal structures that are statistically indistinguishable (e.g. see van der Maas et al. 2006). It is conceivable that common method variance was at least partly responsible for the observation that items tended to reflect a general ASD factor more than specific factors. That is, the general ASD variance could, to some extent, have a methodological root rather than a substantive root and in fact the bi-factor model has sometimes been recommended as a means of partialling out 'nuisance' or 'method' variance common to items in an inventory (Maydeu-Olivares & Coffman, 2006). One mitigating factor is that items of the AQ50 and AQ-S are keyed in both a forward and reverse direction, limiting the amount of common variance due to individual differences in acquiescent response styles. Nonetheless, there remain other sources of common variance of non-substantive origin. These could include, for example, common variance due to implicit theories about autistic behaviours tending to go together which could result in a respondent who holds such an implicit theory answering items more similarly than is merited based on their actual behaviour (e.g. Lahey, Rathouz, Keenan, Stepp, Lober & Hipwell, 2015). Similarly, it could reflect individual differences in the tendency to portray oneself in a socially desirable manner (e.g. Lahey, Applegate, Hakes, Zald, & Rathouz, 2012). Finally, it could reflect context effects of other items whereby responses to previous items are used as a source of information in constructing answers to subsequent items, inflating their similarity (Harrison, McLaughlin & Coalter, 1996). For this reason, it is also important to note that results obtained using the AQ-S could have been affected by the fact that the 28 items were completed in the context of the full 50-item AQ, not in isolation. An important future direction will be to determine why such a strong general factor is found in the AQ50 and AQ-S and, in particular, the extent to which it reflects various influences such as measurement artefacts, sets of shared etiological factors, and local interactions between different symptoms.

Finally, it should be noted that the strength of the general factor in an inventory is inexorably linked to the breadth of behaviours and symptoms covered by the set of items. A strong general factor will tend to be in evidence when items are all very similar to one another (e.g. when some items are mere paraphrases) and measure a narrow construct. The AQ50 and AQ-S are not exhaustive in the features of ASD that they cover and they focus on trait-like behaviours, however, within this they appear to be reasonably diverse in content and don't include many obviously highly redundant items that could inflate item inter-correlations. In terms of the most important areas not represented in the AQ, items focussing on stereotyped behaviours and other features commonly associated with individuals of a low level of functioning are generally absent. Their exclusion is a result of the deliberate targeting of the AQ and AQ-S to individuals of normal intellectual functioning (Baron-Cohen et al. 2001); however, it does limit the relevance of the inventories to the large proportion of individuals with ASD who would be classified as low functioning.

Conclusion

With the exception of the Attention to Details and Numbers/patterns factor of the AQ50 and AQ-S respectively, the items of both inventories appear to primarily reflect a general ASD attribute rather than specific symptoms. This suggests that caution is due when attempting to estimate the relation of a specific symptom with some external criterion which may, as a result of this inventory-wide shared variance, be confounded. In some circumstances, when an association with a specific symptom is of interest, a bi-factor measurement model can be used to separate out general and specific factor variance.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th edition).* Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: *American Psychiatric Association*.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5-17.

Bauman, M. L. (2010). Medical comorbidities in autism: challenges to diagnosis and treatment. *Neurotherapeutics*, *7*, 320-327.

Ben-Sasson, A., Hen, L., Fluss, R., Cermak, S. A., Engel-Yeger, B., & Gal, E. (2009). A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*,*39*, 1-11.

Booth, T., Murray, A. L., McKenzie, K., Kuenssberg, R., O'Donnell, M., & Burnett, H. (2013). Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *Journal of Autism and Developmental Disorders*,*43*, 2997-3000.

Braeken, J. (2011). A boundary mixture approach to violations of conditional independence. *Psychometrika*, *76*, 57-76

Buck, T. R., Viskochil, J., Farley, M., Coon, H., McMahon, W. M., Morgan, J., & Bilder, D. A. (2014). Psychiatric comorbidity and medication use in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*,*44*, 3063-3071.

Burke, J. D., Boylan, K., Rowe, R., Duku, E., Stepp, S. D., Hipwell, A. E., & Waldman, I. D. (2014). Identifying the irritability dimension of ODD: Application of a modified bifactor model across five large community samples of children. *Journal of Abnormal Psychology*, *123*, 841-851

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor one general psychopathology factor in the structure of psychiatric disorders?. *Clinical Psychological Science*,*2*, 119-137.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*, 403-424.

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, *13*, 354-378.

Gadow, K. D., & Sprafkin, J. (2000). *Early Childhood Inventory-4 screening manual.* Stony Brook, NY: Checkmate Plus.

Gadow, K. D., & Sprafkin, J. (2002). *Child Symptom Inventory-4 screening and norms manual*. Stony Brook, NY: Checkmate Plus.

Gillberg, C., Gillberg, C., Råstam, M., & Wentz, E. (2001). The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI): a preliminary study of a new structured clinical interview. *Autism*, *5*, 57-66.

Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain and cognition*,*61*, 25-39.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*, 5-25.

Harrison, D. A., McLaughlin, M. E., & Coalter, T. M. (1996). Context, cognition, and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes*, *68*, 246-261.

Hoekstra, R. A., Vinkhuyzen, A. A., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron- Cohen, S., … & van der Sluis, S. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-Short).*Journal of autism and developmental disorders*, *41*, 589-596.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55.

Johnson, S. A., Filliter, J. H., & Murphy, R. R. (2009). Discrepancies between self- and parent-perceptions of autistic traits and empathy in high functioning children and adolescents on the autism spectrum. *Journal of Autism and Developmental Disorders*, *39*, 1706-1714.

Kuenssberg, R., Murray, A. L., Booth, T., & McKenzie, K. (2014). Structural validation of the abridged Autism Spectrum Quotient–Short Form in a clinical sample of people with autism spectrum disorders. *Autism*, *18*, 69-75.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood?. *Journal of Abnormal Psychology*, *121*, 971-

Lahey, B. B., Rathouz, P. J., Keenan, K., Stepp, S. D., Loeber, R., & Hipwell, A. E. (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry*, *56*, 415-422.

Lecavalier, L., Gadow, K. D., DeVincent, C. J., Houts, C., & Edwards, M. C. (2009). Deconstructing the PDD clinical phenotype: internal validity of the DSM-IV. *Journal of Child Psychology and Psychiatry*, *50*, 1246-1254.

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders*, *35*, 445-460.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659-685.

Mandy, W., Charman, T., Puura, K., & Skuse, D. (2014). Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: Evidence from Finnish and UK samples. *Autism*, *18*, 45-54.

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in developmental disabilities*, *30*, 1107-1114.

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, *11*, 344.

McDonald, R. P. (1999). Test theory: A unified treatment. *Mahwah, NJ: LEA*.

Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014). Are autistic traits measured equivalently in individuals with and without an autism spectrum disorder? An invariance analysis of the Autism Spectrum Quotient Short Form. *Journal of Autism and Developmental Disorders*, *44*, 55-64.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*, 407-422.

Murray, A. L., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014). Are we under- estimating the association between autism symptoms?: The importance of considering simultaneous selection when using samples of individuals who meet diagnostic criteria for an autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 1-10.

Muthén, L. K., & Muthén, B. O. (1998-2013). Mplus U'er's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Posserud, M. B., Breivik, K., Gillberg, C., & Lundervold, A. J. (2013). ASSERT–The Autism Symptom SElf-ReporT for adolescents and adults: Bifactor analysis and validation in a large adolescent population. *Research in Developmental Disabilities*, *34*, 4495-4503.

Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667-696.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*, 129-140.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544-559.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19-31.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*, 57-74.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145-154.

Russell-Smith, S. N., Maybery, M. T., & Bayliss, D. M. (2011). Relationships between autistic-like and schizotypy traits: An analysis using the Autism Spectrum Quotient and Oxford-Liverpool Inventory of Feelings and Experiences. *Personality and Individual Differences*, *51*, 128-132.

Schermelleh-Engel K., Moosbrugger H. & Muller H. (2003). Evaluating the fit of structural equation models: tof significance and descriptive goodness-of-fit measures. *Methods of Psychological Research,* 8, 23-74.

Simms, L. J., Grös, D. F., Watson, D., & O'Hara, M. W. (2008). Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depression and Anxiety*, *25*, E34-E46.

Snow, A. V., Lecavalier, L., & Houts, C. (2009). The structure of the Autism Diagnostic Interview-Revised: diagnostic and phenotypic implications. *Journal of Child Psychology and Psychiatry*, *50*, 734-742.

Stewart, M. E., & Austin, E. J. (2009). The structure of the Autism-Spectrum Quotient (AQ): Evidence from a student sample in Scotland. *Personality and Individual Differences*, *47*, 224-228.

Van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*,*113*, 842.

Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (2013). Construct validity of the WISC–IVUK with a large referred Irish sample. *International Journal of School & Educational Psychology*, *1*, 102-111.

Watson, D.’ O'Hara, M. W., Simms, L. J., Kotov, R., Chmielewski, M., McDade-Montez, E. A., ... & Stuart, S. (2007). Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS). *Psychological assessment*, *19*, 253.

Wechsler, D. (2003). *Wechsler intelligence scale for children (4th ed.).* San Antonio, TX: The Psychological Corporation.

Williams, D. M., & Bowler, D. M. (2014). Autism spectrum disorder: Fractionable or coherent?. *Autism*, *18*, 2-5.

Willoughby, M. T., Blanton, Z. E., & Family Life Project Investigators. (2015). Replication and External Validation of a Bi-Factor Parameterization of Attention Deficit/Hyperactivity Symptomatology. *Journal of Clinical Child & Adolescent Psychology*, *44*, 68-79.

Published as: Murray, A. L., McKenzie, K., Kuenssberg, R., & Booth, T. (2015). Do the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotient Short Form (AQ-S) Primarily Reflect General ASD Traits or Specific ASD Traits? A Bi-Factor Analysis. *Assessment*, 1073191115611230.

Table 1: Standardized factor loadings of the bi-factor CFA (WLSMV) of the AQ50 (n=562)

Item	Factor Loadings							Explained Variance	
	Abbreviate item content	General	Social Skills	Attention Switching	Attention to Detail	Comm.	Imagin	Total	General to total ratio
AQ1	Prefer to do things on own	.46	.43					.40	.54
AQ11	Find social situations easy	.87	.27					.83	.91
AQ13	Prefer library to party	.59	.46					.56	.62
AQ15	Drawn to people over things	.74	.20					.59	.94
AQ22	Hard to make new friends	.81	.10					.67	.99
AQ36	Can infer thoughts/feelings from faces	.77	-.25					.65	.91
AQ44	Enjoy social occasions	.80	.51					.89	.71
AQ45	Difficulty inferring intentions	.79	-.27					.70	.89
AQ47	Enjoy meeting new people	.77	.41					.77	.78
AQ48	Good diplomat	.64	-.10					.42	.98
AQ2	Prefer to do things same way	.56		.46				.53	.60
AQ4	Get so absorbed	.55		.14				.32	.94
AQ10	Keep track of several conversations	.70		-.11				.50	.98
AQ16	Very strong interests	.46		.33				.32	.66
AQ25	Not upset by disruption of daily routine	.53		.47				.50	.55
AQ32	Multi-tasking easy	.62		**-.06**				.38	.99
AQ34	Enjoy doing things spontaneously	.70		.26				.56	.88
AQ37	Can switch back after interruption	.60		**.04**				.36	1.00
AQ43	Like to plan activities carefully	.51		.39				.42	.63
AQ46	Anxiety in new situations	.66		.22				.48	.90
AQ5	Notice small sounds	.28			.29			.16	.48

Published as: Murray, A. L., McKenzie, K., Kuenssberg, R., & Booth, T. (2015). Do the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotient Short Form (AQ-S) Primarily Reflect General ASD Traits or Specific ASD Traits? A Bi-Factor Analysis. *Assessment*, 1073191115611230.

AQ6	Notice information such as number plates	.30			.63			.49	.19
AQ9	Fascinated by dates	.29			.60			.45	.19
AQ12	Notice details others don't	.23			.60			.41	.13
AQ19	Fascinated by numbers	.30			.61			.47	.19
AQ23	Notice patterns all the time	.33			.66			.55	.20
AQ28	Concentrate on whole picture over details	.58			.24			.39	.86
AQ29	Not good at remembering phone numbers	**.07**			.48			.24	.02
AQ30	Don't notice small changes	-.11			.29			.10	.12
AQ49	Not good at remembering dates of birth	**-.05**			.44			.19	.02
AQ7	Inadvertently impolite	.63				.42		.57	.69
AQ17	Enjoy social chit-chat	.78				-.37		.74	.81
AQ18	Difficult for others to get word in edgeways	.19				.48		.27	.13
AQ26	Don't know how to maintain conversation	.82				-.07		.68	.99
AQ27	Easy to read between the lines	.73				.17		.55	.95
AQ31	Can't tell if someone is bored listening	.62				.20		.42	.91
AQ33	Can turn-take talking on phone	.59				.18		.38	.91
AQ35	Last to understand point of joke	.59				.27		.42	.82
AQ38	Good at social chit-chat	.88				-.38		.92	.85
AQ39	Go on about the same thing	.61				.60		.73	.51
AQ3	Hard to create picture in mind	.38					.49	.38	.38
AQ8	Easy to imagine characters in a story	.53					.50	.52	.53
AQ14	Making up stories is easy	.24					.54	.35	.17

AQ20	Working out character intentions in story	.62	.25	.45	.86
AQ21	Don't enjoy fiction	.37	.38	.28	.49
AQ24	Prefer theatre to museum	.42	**.04**	.18	.99
AQ40	Enjoyed pretend play as child	.55	.36	.43	.71
AQ41	Collect information about categories of things	.61	**-.07**	.38	.99
AQ42	Difficult to imagine being someone else	.62	.19	.41	.91
AQ50	Easy to play pretending games with children	.59	.40	.51	.69

Note: Comm. = Communication; Imagin. = Imagination. Values in boldface are non-significant at p<.05.

Published as: Murray, A. L., McKenzie, K., Kuenssberg, R., & Booth, T. (2015). Do the Autism Spectrum Quotient (AQ) and Autism Spectrum Quotient Short Form (AQ-S) Primarily Reflect General ASD Traits or Specific ASD Traits? A Bi-Factor Analysis. *Assessment*, 1073191115611230.

Table 2: ECV, Omega Total, Omega Hierarchical, Beta and PRMSE for the AQ50 and AQS

	ECV	ω_T	ω_h	β	PRMSE Total Score	PRMSE Modified Total score	PRMSE Subscale Score
AQ50	.70						
General		.97	.91	.83	-	-	-
(excluding Attention to Details items)		(.97)	(.80)	(.89)			
Social Skills		.94	.05	.83	.84	.89	.91
Attention Switching		.87	.10	.74	.85	.85	.84
Attention to Details		.81	.67	.59	.33	-	.75
Communication		.91	.05	.73	.88	.91	.86
Imagination		.85	.24	.71	.78	.80	.81
AQS	.63						
General		.96	.86	.83	-	-	-
(excluding Number/Patterns items)		(.96)	(.71)	(.87)			
Social Skills		.93	.27	.84	.79	.85	.90
Routine		.82	.18	.72	.77	.77	.76
Switching		.82	.19	.71	.73	.73	.73
Imagination		.87	.21	.71	.76	.76	.82
Numbers/Patterns		.83	.62	.70	.38	-	.77

Note. For the AQ50, the modified total score excludes the Attention to Detail items and for the AQ-S it excludes the Numbers/patterns items

Table 3: Standardized factor loadings of the bi-factor CFA (WLSMV) of the AQ-S (n=562)

Item		Factor loadings						Explained variance	
		General	Social Skills	Routine	Switching	Imagin	Numbers/ patterns	Total	General to total ratio
AQ1	Prefer to do things on own	.40	.49					.40	.40
AQ11	Find social situations easy	.81	.42					.83	.79
AQ13	Prefer library to party	.51	.54					.55	.47
AQ15	Drawn to people over things	.72	.30					.61	.86
AQ22	Hard to make new friends	.78	.25					.66	.91
AQ44	Enjoy social occasions	.69	.64					.88	.54
AQ47	Enjoy meeting new people	.70	.53					.78	.63
AQ46	Anxiety in new situations	.66	.17	.20				.50	.87
AQ2	Prefer to do things same way	.59		.37				.48	.72
AQ25	Not upset by disruption of daily routine	.55		.52				.57	.54
AQ34	Enjoy doing things spontaneously	.74		.25				.61	.89
AQ4	Get so absorbed	.53			.17			.31	.91
AQ10	Keep track of several conversations	.72			.16			.54	.95
AQ32	Multi-tasking easy	.60			.80			.99	.36
AQ37	Can switch back after interruption	.61			.24			.43	.86
AQ3	Hard to create picture in mind	.38				.57		.47	.31

AQ8	Easy to imagine characters in a story	.53				.51		.54	.53
AQ14	Making up stories is easy	.26				.52		.34	.20
AQ20	Working out character intentions in story	.61				.24		.43	.87
AQ36	Can infer thoughts/feelings from faces	.75				.13		.58	.97
AQ42	Difficult to imagine being someone else	.62				.18		.42	.92
AQ45	Difficulty inferring intentions	.77				.13		.62	.97
AQ50	Easy to play pretending games with children	.61				.27		.45	.84
AQ6	Notice information such as number plates	.29					.62	.46	.18
AQ9	Fascinated by dates	.27					.69	.55	.13
AQ19	Fascinated by numbers	.28					.72	.60	.13
AQ23	Notice patterns all the time	.31					.64	.50	.19
AQ41	Collect information about categories of things	.58					.32	.44	.77

Note: Imagin = Imagination. All loadings are significant at p<.05.

Table 4: Standardized coefficients for the effect of sex on specific factors from the correlated traits and bi-factor models for the AQ-50 and AQ-S (n=562)

	Correlated Traits (95%CI)	**Bi-factor (95%CI)**
AQ50		
Social Skills	-.28 (-.36 to -.20)	-.29 (-.42 to -.15)
Attention Switching	-0.34 (-.42 to -.26)	-0.63 (-.73 to -.52)
Attention Detail	-0.19 (-.28 to -.11)	-0.15 (-.24 to -.06)
Communication	-0.33 (-.40 to -.25)	-0.33 (-.43 to -.24)
Imagination	-0.36 (-.44 to -.28)	-0.45 (-.56 to -.34)
AQ-S		
Social Skills	-0.25 (-.33 to -.17)	-0.36 (-.49 to -.24)
Routine	-0.25 (-.34 to -.16)	-0.45 (-.60 to -.29)
Switching	-0.40 (-.48 to -.32)	-0.64 (-.75 to -.54)
Imagination	-0.35 (-.43 to -.27)	-0.51 (-.62 to -.40)
Numbers/ Patterns	-0.27 (-.36 to -.19)	-0.26 (-.35 to -.16)